Challenges and examples of in-situ memory content extraction techniques

Franck Courbon
Department of Computer Science and Technology
University of Cambridge
Cambridge, UK
franck.courbon@cl.cam.ac.uk

***Abstract*—We present embedded devices memory elements - from core registers to off chip-use, type and architecture before summarising their features regarding extraction techniques at scale. We list recent and on-going attack platform methodologies prior analysing their pros and cons. Particularly of importance, we address combined attack approaches, signal processing techniques and the challenges of low cost extraction methodologies. Above all, we characterise beam-based extraction techniques, starting from sample preparation before concluding on in-situ memory content extraction limits and countermeasures.**

***Index Terms*—Memory content extraction, combined attack, low cost approach, beam-based techniques**

INTRODUCTION

Power efficiency, area, density, throughput, IP price are common concerns when choosing storage elements for the required application. Mainly due to integrated circuits field of application (military, payment, healthcare, cyberphysical system and other connected objects), hardware security is taking a step forward at the design stage [1] to cover various physical attacks [2]. While an algorithm is proven mathematically sure, its hardware implementation made of transistors is subject to perturbation (glitch, laser) or side-channel leakage (power, time). Beside those attacks, there are also requirements in terms of counterfeit protection and malicious circuits. EDA tools security related capabilities are currently being developed to make embedded system secure by design and avoid impossible (hardware fix for next device versions only) or costly fix (performance downgrade) situations.

When designing a new secure device and protecting it against physical attacks, the first developed countermeasures aim to eradicate product reverse engineering. Physical attacks focus on the digital logic (including the core(s)) and memories. Some encryption/decryption blocks are implemented using a certain number of interconnected standard cells. They can be camouflaged at various hardware level (silicon, drain/source/gate, vias) [3]. Memory elements design have evolved from visually distinguishable states (ROM) and metal/poly fuses to electric charge/polarization states and anti-fuse technologies. Countermeasures against in-situ extraction techniques range from address scrambling at the hardware level to software encryption/key derivation schemes.

Dr. Franck Courbon is an Early Career Fellow funded by the Isaac Newton Trust and the Leverhulme Trust Early Career Fellowship (ECF-2017-606)

Design stage security principles are mainly based on security models that cover known vulnerabilities only. The purpose of this work is to analyze current attacker capabilities (with low cost means) and name some of the remaining challenges for in-situ content extraction. We particularly point out combined approaches and this article contributes to analyse if low hanging fruits (best attacks in terms of cost, time and efficiency) are hardware located as in recent attacks [4].

I. EMBEDDED DEVICE MEMORY

A. Role, application and type

Embedded devices memory elements range from single bit registers (flip-flops) within the devices core(s), on-chip non volatile (ROM/EEPROM/FLASH/eFuses) or volatile memory (SRAM) to off-chip volatile memory (as DRAM in mobile phone Package on Package (PoP)). Depending on their density and time access, they are used as/for working registers, keys storage, boot loaders, protection against firmware downgrade, small variables and large data. Hardware extraction techniques can be used as the last possible chance to extract information from a secure device (security), from a damaged device (forensic) or from a not functional/not completed device (manufacturing). Security wise, in-situ memory content extraction aims to access critical root of trust information or sensitive algorithm/data/keys to facilitate combined physical attacks or enables a software attack. Adding to the above list of standard memory types, there are also emerging technologies mainly enabling very low-power designs such as Phase Change Memory (PCM), Ferroelectric RAM (FRAM), Magnetic RAM (MRAM), Oxide-based RAM (OxRAM).

B. Structure, materials and vertical position

The standard structure of a 45nm and above (45nm corresponding to the transistor gates width) integrated circuit (CMOS based) includes a Silicon substrate, doped areas (transistors' drain and source), poly-Silicon (transistors' gate) and a stack of metal layers and dielectrics interconnected by vias.

Depending on the package type you can then find some various elements such as the following ones for a 45nm FCBGA: Passivation: $Si_3N_4/SiO_2/Si_3N_4$ (0.6/0.1/0.6μm), Polyimide (5μm), Die bumps, PCB substrate and Copper balls. While the thickness of each layer of the device starts only from $0.2\mu m$, the area typically covers $6000\mu m$ by $8000\mu m$ and

depends on the complexity of the device (evaluated in kGE (Gate Equivalent)). The Si. substrate is $650\mu m$ thick.

At the time being, memory information is either stored at the top of the device (e.g. MRAM/OxRAM (large elements), BEOL: Back-End-Of-Line) or more commonly below the first metal layer (e.g. FRAM, FEOL: Front-End-Of-Line). For any of the given memory type, the in-situ extraction technique has to be evaluated according to the following criteria:

- Is a direct access to the storage layer required?
- Is the required sample preparation expensive/hard?
- Can the technique scale in terms of speed, area?
- What are the techniques limits (efficiency, process)?
- Can the raw extraction technique be combined (e.g. signal processing, error correction, multiple samples)?

C. In-situ extraction techniques

Hardware in-situ extraction techniques may target embedded devices for which security is essential, the tools behind such extraction can thus be costly such as the following ones:

- Photon-emission analysis (SRAM) [5].
- Focused Ion Beam (FIB) extract. (and mod.) (Fuses) [6].
- X-ray (ptychography) memory modification (SRAM) [7].

However, those setup are either expensive, not open source or not demonstrated on recent technology node devices. Similarly to active circuit electromagnetism observation, Xray methods are very interesting due to their application through the device package (non invasive). At the moment, the demonstrated technique requires a synchrotron, time and expertise but multiple electron sources based Xray characterizion re being developed [8]. Also, in the field, there are current development in laser platform with thick substrate capability or multiple beam spots.

D. Towards low-cost and combined techniques

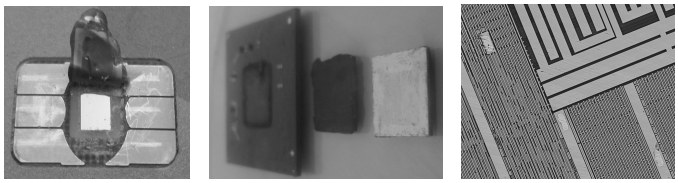
We also see the development of low cost laser setups, that follows low cost approach used in other field such a powerful paper based microscope [9] or motorized stage [10]. Having low-cost approaches directly impact attacks classification. We also see the advances in signal processing techniques and for instance the efforts in the field of biology that can be, brought and enhanced for hardware security applications. Combined techniques from hardware security, forensic analysis and failure analysis (including debug and test) communities are also an interesting approach.

In the next section, we bring together methods to access the layer of interest where the memory information is, no matter the type of package and the type of memory.

II. Developping low cost memory layer access

A. Die frontside and backside opening

For in-situ memory content extraction requiring a direct access to the layer where the information is stored (for instance at the transistors gate level), one also has to remove the component from its package. Packages vary from chip modules, QFN like package with multiple external pins and Ball Grid Arrays. Nowadays, for cost reduction, most of the chip in modules or BGA are said flip-chip. Opening the device from the front of the package would give access to the back of the chip itself (the substrate), Fig. 1a. The die extraction is realized with a sharp knife (smart card module) only or simply combined with a hot plate (FCBGA) to first remove the die from the PCB and then grind connecting Cu balls. A complete wet etching backside removal while protecting edges would also be of interest for the community.



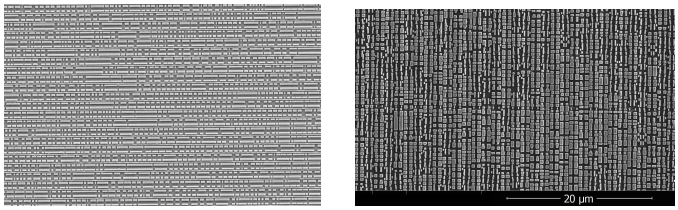
(a) Smart card module (b) FCBGA module (c) SEM top metal

Fig. 1: Embedded system depackaging

Depending on the application, it is already possible to look through optically transparent material (polyimide and passivation) for the FCBGA device, right image of Fig. 1. On this particular image, one can notice three gray rectangles corresponding to three balls (removed) localization. Various structures can be seen enabling the identification/localisation of certain type of memory close to the top metal layer.

B. Destructive access to transistor's active region (frontside)

On Fig. 2, one can identify standard differences between the smart card module and the FCBGA. Those images are obtained after wet etching (HF acid) where only active regions remain. For the FCGBA, the technology process is smaller (45 vs 90nm), there are usually no well taps and countermeasures are not present against invasive attacks (reverse engineering). The technique only requires HF acid (few minutes bath), acetone, ultrasonic rinsing bath and a nitrogen gun to rinse the circuit. The sample preparation is easy, fast and efficient no matter the size of the circuit and the technology node.



(a) A smart card IC (b) A processor

Fig. 2: Frontside Scanning Electron Microscopy (active layer)

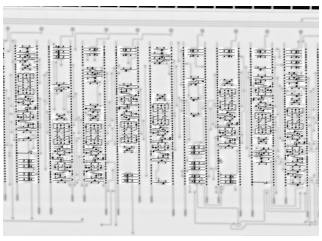
A small modification on the methodology could be brought to conserve Tungsten plugs (M1 contacts) if necessary. For instance it could enable with a successive electron beam imaging the extraction of ROM contents if encoded at this

layer or more importantly, provide access to other type of memory present below the first metal layer.

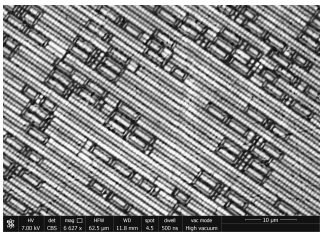
C. Invasive access to transistor's active region (backside)

Another approach is to let the metal layers on the circuit and apply low-cost methods to remove all Silicon substrate to access transistors layer from the backside. Manual sandpaper polishing (with various grit) combining with Colloidal Silica is a possibility while polishing machine can also be used. A final selective Si/Sio2 etch (Choline hydroxide) permit to remove the remaining Silicon substrate.

However, the task is complicated for large package. Indeed, there is a die curvature due to the packaging process. While the goal is to only have few μm of Silicon remaining on the chip, there can be tens of μm difference between the circuit center and its sides. One solution is adaptive CNC milling. Also, one can note that some expensive tools (100kEuros) including an interferometer for local measurement can remove Silicon down to few μm over the whole area of a circuit. Backside access is very promising as the device could remain functionnal (if some tens of nm are left on the circuit), left image on Fig. 3. It could also enable to analyze the logic located next to a memory or find where memory elements such as flip-flops are located on the chip and how memory related elements are connected together, right image on Fig. 3. This image is obtained with Backscattered Electrons (BSE) detector giving more sub-surface information.



(a) Seeing through few nm of Si.



(b) BSE image example

Fig. 3: Backside Scanning Electron Microscopy (active layer)

After presenting low cost backside and frontside access, we introduce some of our current in-situ memory memory content extraction techniques. Those extraction techniques are low-cost, large area compliant and fast.

III. Extraction/Characterization techniques

A. Various memory elements

Table I presents the different memory types that can be found in a circuit. While ROM, doped ROM and Fuse elements have been widely reverse engineered using sample preparation and imaging capabilities, the next step is to further analyse their respective replacing technologies; EEPROM/FLASH and Antifuse. Particularly of interest are antifuse technologies due to manufacturers security statement and emerging ChipLock technology [11]. An additional step would be to assess emerging nonvolatile memory in-situ extraction, the main particularity is to access where the information is stored. In-situ single extraction from DRAM (volatile and mainly data) is not practical. DRAM would instead require techniques developed in the forensic analysis field (not cell level extraction).

TABLE I: Memory type analysis

Storage Element	Features		
	Storage type	***Localization***	***Volatile?***
ROM	OTP by material	FEOL or BEOL	No
Doped ROM	OTP by doping	FEOL	No
EEPROM	Electric charge	FEOL	No
FLASH	Electric charge	FEOL	No
SRAM	Memory based	FEOL	Yes
DRAM	Capacitance	FEOL	Yes
FRAM	Polarization	FEOL	No
PCM	Material Phase	FEOL	No
MRAM	Magnetic orientation	BEOL	No
CBRAM	Electric resistance	BEOL	No
OxRRAM	Electric resistance	BEOL	No
Flip-flop	Memory based (logic)	FEOL	No
Fuse	Open/Short	FEOL or BEOL	No
Antifuse	Oxide breakdown	FEOL	No
ChipLock	Multi-layer pattern	FEOL & BEOL	No

B. Electric charge characterization using Scanning Electron Microscope

SEM based techniques (without nanoprobes) benefit from the capability to define an area (with a certain magnification and overlay) that will be automatically acquired without operator interaction (unlike Atomic Force Microscopy (AFM) approaches). One of the application is EEPROM/Flash content extraction [12] and it benefits from both SEM capabilities and offline image processing techniques. We show on Fig. 4 the capability to perform such memory content extraction over two successive acquired EEPROM area.

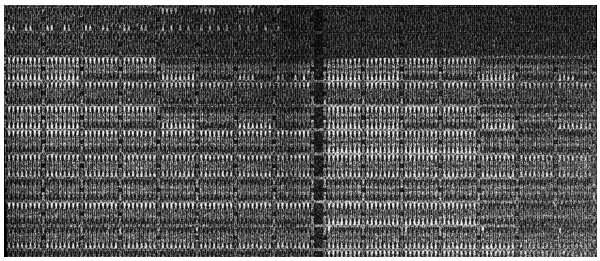
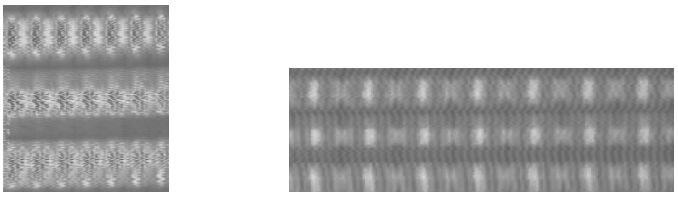


Fig. 4: Multi-area EEPROM extraction using SEM

C. Laser based reverse engineering

A laser scanning technique (with no high spatial and temporal precision setup) has been tested over SRAM memory cells, Fig. 5a. The device is ON and the power line of the device is monitored. A continuous laser beam interfers with each element of the memory and the power signal is retrieved and plotted. One can notice that no matter the timing approximation between each line of power analysis, standard open source technique can permit to correctly align each line to each other. The drawback of laser based reverse engineering is its impact on the device's power line that can be easily

detected (the device has to be in "ON" for memories such as SRAM), thus limiting the field of application. However, this technique could be further extended to some targets in the IoT regime not particularly (yet?) aware of such hardware flaw.



(a) Raw scan (b) Re-aligned scan

Fig. 5: SRAM cells Laser Scanning Microscopy (LSM)

D. Flip-flop encryption/state extraction

While full reverse engineering remain too costly in terms of skills, equipment and time, partial reverse engineering is an interesting first step for combined attack, memory state extraction and hardware crypto/encryption accelerator analysis. We developed a machine learning based analysis of pattern present at the transistors active layer, Fig. 6. We are able to draw statistics within a single chip, from a chip to another and so on. Experiments and classifications are still undergoing regarding memory related logic circuitry. Machine learning can also enable to resolve ultrafast electron microscope scans or low resolution optical images (also applying a specific solution over the device can help increasing the image resolution). It particularly becomes of interest following the advances of open source hardware based on RISC-V Instruction Set Architecture (ISA). For the SoC device under test, parameters were the following:

1) Time per pixel (dwell): $1\mu s$, res.: 3072 * 2048 pixels
2) Complete chip size: 142k * 159k pixels (9.3 *10.4 mm)
3) Acquis. time: 8.5 hours (Overlap: 10%), Align.: 5 Min.

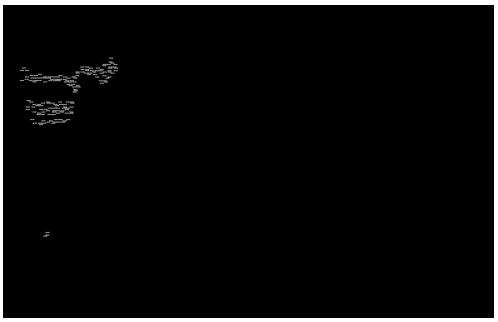


Fig. 6: Chip area subset single cell localization extraction

Recent advances in multi-electron beam (91 beams) reduce such time acquisition to less than 10 minutes. We use offline tools for image alignment after obtaining artifacts using closed source proprietary SEM software.

E. Combining multi-field approaches

Remaining challenges vary from the application on recent products of non-volatile memory extraction, eFuse in-situ extraction, large scale caracterization of hardware based encryption techniques and the use of other failure analysis/chip debugging tools. There is also a strong interest for low-cost sample preparation (chemistry and polishing) allowing any CMOS circuit preparation. Similarly, sample preparation on devices with an insulator structure (BOX) above the substrate could be characterized. Other type of failure analysis, chip manufacturing tools and caracterization techniques need to be studied. For instance, the state of the art does not cover if standard laboratory available laser based technique (raman, ellipsometer) or scanning near- field optical microscopy (SNOM) are possible in-situ memory content extraction tools. At last, presented techniques have mostly been demonstrated on a subset of the device under test, there is a demand for global application and preferably on current process nodes/devices to analyze techniques pros and cons and define necessary countermeasures. While state of the art countermeasures might lock most exploits for a single type of attack (side channel, fault), another interesting path is combined approaches. For instance, combining physical attacks to make debug interfaces accessible could be beneficial for an attacker. There is also a particular interest with techniques coming from the forensic analysis world (e.g. chip off techniques).

CONCLUSION

This paper gives a study of the various type of memory elements present in any embedded devices and their localization in the hardware stack. We present some of our latest advances in low cost, fast and efficient approaches combining sample preparation and beam based (electron, laser) techniques for memory content extraction. Some of the numerous remaining challenges are introduced and we particularly point out the interest of combined approaches. Moving to an era of uniqueness and specialization (and open source hardware) for more speed and less power consumption push even further the questions related to hardware security and in-situ memory content extraction techniques.

REFERENCES

[1] H. Khattri, N. K. V Mangipudi, S. Mandujano, HSDL: A Security Development Lifecycle for hardware technologies, 2012.
[2] S. Skorobogatov, Semi-invasive attacks, A new approach to hardware security analysis, 2005.
[3] R. Cocchi, Camouflage circuitry and programmable cells to secure semiconductor designs during manufacturing, NAECON, 2015.
[4] https://news.softpedia.com/news/nvidia-tegra-x1-coldboot-vulnerabilitylets-anyone-hack-a-nintendo-switch-520811.shtml
[5] F. Stellari, P. Song, M. Villalobos, J. Sylvestri, Revealing SRAM memory content using spontaneous photon emission, VTS 2016.
[6] C. Tarnovsky, Security Failures In Secure Devices, 2008.
[7] S. Anceau, P. Bleuet, J. Cledi ere, L. Maingault, J-L Rainard and R ' emi Tucoulou, Nanofocused X-ray Beam to Reprogram Secure Circuits, Cryptographic Hardware and Embedded Systems (CHES), 2017.
[8] R. Lanza, Nanoscale X-Ray Tomosynthesis for Rapid Assessment of IC Dice, AIDA-2020 Meeting, 2018.
[9] J. S. Cybulski, J. Clements, M. Prakash, Foldscope: Origami-Based Paper Microscope, 2014.
[10] R. Campbell, R. Eifert, G. Turner, Openstage: A Low-Cost Motorized Microscope Stage with Sub-Micron Positioning Accuracy, 2014.
[11] Multibeam ChipLock, http://www.multibeamcorp.com/applications.html
[12] F. Courbon , S. Skorobogatov, C. Woods, Reverse engineering Flash EEPROM memories using scanning electron, CARDIS, 2017